

REPORT



과목명		비모수통계자료분석
담당교수		최영훈 교수님
학과		응용통계학과
학년		3학년
학번		201452024
이름		박상희
제출일		2018년 11월 21일

연습문제 3

한 소비자 단체에서는 어린이 장난감에 많이 사용하는 4종류의 건전지 평균수명간의 차이를 비교하고자 한다. 이를 위해 4종류의 건전지로부터 임의의 확률표본을 추출하여 연속사용으로 인한 수명(단위 : 분)을 측정한 결과가 아래에 주어져있다. 건전지 종류 간의 상대적 평균수명의 우위가 존재하는지를 밝혀라.

건전지1	건전지2	건전지3	건전지4
48(27)	49(29.5)	41(3)	55(34)
46(20)	52(32)	38(1)	58(36)
43(8)	45(14)	45(14)	45(14)
46(20)	44(10)	45(14)	46(20)
47(24)	48(27)	43(8)	48(27)
42(5.5)	45(14)	46(20)	50(31)
45(14)	47(24)	41(3)	56(35)
49(29.5)	46(20)	43(8)	47(24)
45(14)		41(3)	53(33)
		42(5.5)	
$n_1 = 9$ $R_{1.} = 162$	$n_2 = 8$ $R_{2.} = 170.5$	$n_3 = 10$ $R_{3.} = 79.5$	$n_4 = 9$ $R_{4.} = 254$

① 귀무가설 vs 대립가설

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1 : \text{Not } H_0$$

② Kruskal-Wallis 검정통계량

$$T = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \sim \chi_{k-1}^2$$

$$T = \frac{12}{36(36+1)} \left(\frac{162^2}{9} + \frac{170.5^2}{8} + \frac{79.5^2}{10} + \frac{254^2}{9} \right) - 3(36+1) \sim \chi_{4-1}^2$$

$$T = 18.28 \sim \chi_3^2$$

③ 유의수준 5% 하에서의 통계적 검정

$T = 18.28 > \chi_{3,0.95}^2 = 7.815$ 이므로 귀무가설을 기각한다. 따라서 전지 종류 간의 상대적 평균수명의 차이가 존재한다고 볼 수 있다.

연습문제 8

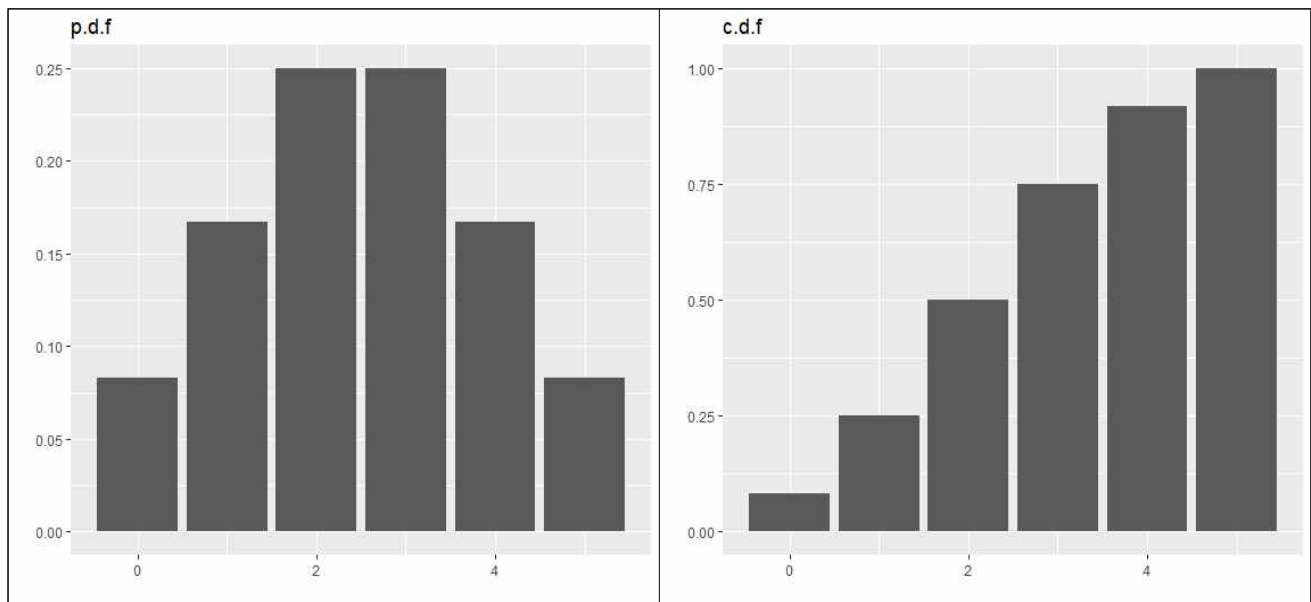
귀무가설 $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ 하에서 $k = 3$ 및 $n_1 = n_2 = 1, n_3 = 2$ 일 때의 Jonckheere 검정통계량 $J = \sum_{i < j}^k U_{ij}$ 의 정확한 분포를

구해 보아라. 이 때 가능한 경우의 수는 $\frac{4!}{1!1!2!} = 12$ 이다.

가능한 경우의 수	$X_{in} < X_{jm}$ 인 쌍의 수
(1) (2) (3,4)	$1 + 2 + 2 = 5$
(1) (3) (2,4)	$1 + 2 + 1 = 4$
(1) (4) (2,3)	$1 + 2 + 0 = 3$
(2) (1) (3,4)	$0 + 2 + 2 = 4$
(2) (3) (1,4)	$1 + 1 + 1 = 3$
(2) (4) (1,3)	$1 + 1 + 0 = 2$
(3) (1) (2,4)	$0 + 1 + 2 = 3$
(3) (2) (1,4)	$0 + 1 + 1 = 2$
(3) (4) (1,2)	$1 + 0 + 0 = 1$
(4) (1) (2,3)	$0 + 0 + 2 = 2$
(4) (2) (1,3)	$0 + 0 + 1 = 1$
(4) (3) (1,2)	$0 + 0 + 0 = 0$

p.d.f	
0	$\frac{1}{12} = 0.083$
1	$\frac{2}{12} = 0.167$
2	$\frac{3}{12} = 0.25$
3	$\frac{3}{12} = 0.25$
4	$\frac{2}{12} = 0.167$
5	$\frac{1}{12} = 0.083$

c.d.f	
0	$\frac{1}{12} = 0.083$
1	$\frac{3}{12} = 0.25$
2	$\frac{6}{12} = 0.5$
3	$\frac{9}{12} = 0.75$
4	$\frac{11}{12} = 0.917$
5	$\frac{12}{12} = 1.0$



연습문제 11

심장질환으로 고생하는 환자 중에서 Group1은 방사선치료를, Group2는 외과수술을, Group3은 약물치료를 실시하였다. 일정한 기간의 치료 후 가 Group별로 추출한 표본의 정산인과 비교한 심장박동비율(%)을 조사한 자료가 아래에 제시되어 있다. 세 모집단 간의 심장박동비율이 $Group2 \leq Group1 \leq Group3$ 의 증가순위인지를 Jonckheere 검정을 이용하여 밝혀라. 이 때의 p값은?

Group	심장박동비율						
Group 1	67	79	57	66	71		
Group 2	42	61	94	76	49	36	
Group 3	101	93	85	83	90	80	82

그룹 재배치

Group	심장박동비율						
Group 2	42	61	94	76	49	36	
Group 1	67	79	57	66	71		
Group 3	101	93	85	83	90	80	82

순서쌍 찾기

i	j	$X_{in} < X_{jm}$ 인 쌍의 수							U_{ij}
2	1	5	4	0	1	5	5		20
2	3	7	7	1	7	7	7		36
1	3	7	7	7	7	7	7		35

Jonckheere 검정통계량

$$T = 20 + 36 + 35 = 91$$

유의수준 5% 하에서 통계적 검정

$T = 91 > X_{0.95} = 74$ 이므로 귀무가설을 기각한다. 따라서 세 모집단 간의 심장박동비율이 $Group2 \leq Group1 \leq Group3$ 의 증가순위라고 볼 수 있다. 또한 p값은 표12에 의하면 $0 < \hat{\alpha} < 0.005$ ($T=85$) 이다.

연습문제 7

고속도로 상에서 새로 선보일 신형 스포츠 자동차의 브레이크 실험 자료를 이용하여 자동차 속력이 증가할수록 제어거리의 분산이 증가하는 경향이 있는지를 밝혀라. 이때의 p값은?

속력	제어거리 (m)									
100km	85	90	77	65	88	94	79	81	98	75
150km	112	118	106	98	83	125	137	105	119	123

<Conover 제곱 순위 검정>

가설

$$H_0 : \text{Var } X \geq \text{Var } Y$$

$$H_1 : \text{Var } X < \text{Var } Y$$

원자료 및 Median 확인

속력	제어거리 (m)										
100km	85	90	77	65	88	94	79	81	98	75	$X_{Median} = 83$
150km	112	118	106	98	83	125	137	105	119	123	$Y_{Median} = 115$

Median과의 차이

속력	제어거리 (m)										
100km	2	7	-6	-18	5	11	-4	-2	15	-8	$X_{Median} = 83$
150km	-3	3	-9	-17	-32	10	22	-10	4	8	$Y_{Median} = 115$

Median과의 차이의 절댓값

속력	제어거리 (m)										
100km	2	7	6	18	5	11	4	2	15	8	$X_{Median} = 83$
150km	3	3	9	17	32	10	22	10	4	8	$Y_{Median} = 115$

Median과의 차이의 절댓값의 순위

속력	제어거리 (m)										
100km	1.5	9	8	18	7	15	5.5	1.5	16	10.5	$X_{Median} = 83$
150km	3.5	3.5	12	17	20	13.5	19	13.5	5.5	10.5	$Y_{Median} = 115$

Median과의 차이의 절댓값의 순위의 제곱합

속력	제어거리 (m)										
100km	2.25	81	64	324	49	225	30.25	2.25	256	110.25	$X_{Median} = 83$
150km	12.25	12.25	144	289	400	182.25	361	182.25	30.25	110.25	$Y_{Median} = 115$

검정통계량

$$T = 2.25 + 81 + 64 + 324 + 49 + 225 + 30.25 + 2.25 + 256 + 110.25 = 1144$$

통계적 검정

통계량 값이 작을수록 X의 분산이 작아지는 경향이 있다. 그러므로 통계량이 작을수록 귀무가설을 기각한다. 따라서 귀무가설을 기각하기 위한 검정은 좌측 검정이다. 따라서 유의수준 5%하에서 $T = 1144 > X_{0.05} = 963$ (표13, $n = m = 10$) 이므로 귀무가설을 채택한다. 자동차 속력이 증가할수록 제어거리의 분산이 증가하는 경향이 있다고 보기 어렵다.

〈 Siegel – Tucky 검정 〉

가설

$$H_0 : Var X \geq Var Y$$
$$H_1 : Var X < Var Y$$

원자료 정렬 및 순위

정렬	65	75	77	79	81	83	85	88	90	94	98	98	105	106	112	118	119	123	125	137
순위	1	4	5	8	9	12	13	16	17	20	19	18	15	14	11	10	7	6	3	2
표본	X	X	X	X	X	Y	X	X	X	X	X	Y	Y	Y	Y	Y	Y	Y	Y	Y

검정통계량

$$T = 1 + 4 + 5 + 8 + 9 + 13 + 16 + 17 + 20 + 19 = 112$$

통계적 검정

통계량 값이 클수록 X의 분산이 작아지는 경향이 있다. 그러므로 통계량이 클수록 귀무가설을 기각한다. 따라서 귀무가설을 기각하기 위한 검정은 우측 검정이다. 따라서 유의수준 5%하에서 $T = 112 < X_{0.95} = 127$ (표9, $n = m = 10$, $2E(T) - X_{0.05} = 210 - 83 = 127$) 이므로 귀무가설을 기각한다. 자동차 속력이 증가할수록 제어거리의 분산이 증가하는 경향이 있다.

유의수준

$$\hat{\alpha} = P(T > 112 | E(T)) = \frac{10(20+1)}{2} = 105$$
 이므로, $0.10 < \hat{\alpha} < 1.0$ 이다.

〈 Ansari – Bradley 검정 〉

가설

$$H_0 : Var X \geq Var Y$$
$$H_1 : Var X < Var Y$$

원자료 정렬 및 순위

정렬	65	75	77	79	81	83	85	88	90	94	98	98	105	106	112	118	119	123	125	137
순위	1	2	3	4	5	6	7	8	9	10	10	9	8	7	6	5	4	3	2	1
표본	X	X	X	X	X	Y	X	X	X	X	X	Y	Y	Y	Y	Y	Y	Y	Y	Y

검정통계량

$$T = 1 + 2 + 3 + 4 + 5 + 7 + 8 + 9 + 10 + 10 = 59$$

통계적 검정

통계량 값이 클수록 X의 분산이 작아지는 경향이 있다. 그러므로 통계량이 클수록 귀무가설을 기각한다. 따라서 귀무가설을 기각하기 위한 검정은 우측 검정이다. 따라서 유의수준 5%하에서 $T = 59 < X_{0.95} = 66$ (표14, $n = m = 10$) 이므로 귀무가설을 채택한다. 자동차 속력이 증가할수록 제어거리의 분산이 증가하는 경향이 있다고 보기 어렵다.

유의수준

$$\hat{\alpha} = P(T > 59)$$
 이므로, $\hat{\alpha} = 3014$ (표14, $n = m = 10$) 이다.